

BAB III

METODOLOGI PENELITIAN

3.1. Objek Penelitian

Objek penelitian pada proposal skripsi ini mahasiswa UMN, banyak mahasiswa UMN ingin cepat lulus atau ingin memperbaiki nilai pada semester antara, untuk program Sarjana adalah jenjang pendidikan akademik yang mempunyai beban studi antara 144 satuan kredit semester (sks) hingga 160 sks, dengan kurikulum 8 semester dan lama program antara 7 sampai dengan 14 semester [29], Program Diploma Tiga (D-3) adalah jenjang pendidikan akademik yang mempunyai beban studi antara 108 hingga 120 satuan kredit semester (sks), dengan kurikulum 6 semester dan lama program antara 6 hingga 10 semester [30].

Proses belajar mengajar di UMN menggunakan metode pembelajaran berbasis pengalaman (*experience-based learning*), para tenaga pengajar mengajarkan pengalamannya sebagai praktisi profesional di bidang yang akan dimasuki mahasiswa dalam tiap program studi pilihan. Metode pembelajaran dilaksanakan melalui berbagai latihan, riset lapangan, praktik kerja, dan *role play* yang disertai penguasaan teori [29] [30].

UMN menyediakan berbagai program studi pilihan yang mengantar para mahasiswa menjadi para pemimpin yang memiliki jiwa

technopreneurship sejati di masa depan. Untuk itu, UMN mempersiapkan para mahasiswa sejak dini dengan menyelenggarakan mata kuliah pengembangan kepribadian yang dapat meningkatkan kemampuan intelegensia dan keahlian dasar para mahasiswa, terutama di bidang penguasaan bahasa asing, penggunaan aplikasi, dan kemampuan berwirausaha berbasis teknologi (*technopreneurship*) [29] [30].

3.2. Data Collection

Data yang digunakan adalah data Kelulusan Mahasiswa Sarjana (S1) yang sudah lulus tahun 2018 Hingga 2020 di Universitas Multimedia Nusantara. data diperoleh dari Biro Informasi Akademik (BIA) Universitas Multimedia Nusantara yang disetujui oleh ketua prodi Sistem Informasi yaitu Ibu Ririn Ikana Desanti, baris data berjumlah 3625 *row* data.

Tabel 3.1. Tabel Atribut, Keterangan, dan Tipe Data

| No | Atribut | Keterangan | Tipe data |
|----|------------------|----------------------------|----------------|
| 1 | NIM | Nomor Induk Mahasiswa | <i>Nominal</i> |
| 2 | NAMA | Nama Mahasiswa | <i>Nominal</i> |
| 3 | ANGKATAN | Angkatan Mahasiswa | Real |
| 4 | PRODI | Prodi Mahasiswa | <i>Nominal</i> |
| 5 | TANGGAL YUDISIUM | Tanggal Yudisium Mahasiswa | Date_time |
| 6 | JENIS KELAMIN | Jenis Kelamin Mahasiswa | <i>Nominal</i> |
| 7 | TEMPAT LAHIR | Tempat Lahir Mahasiswa | <i>Nominal</i> |

| No | Attribut | Keterangan | Tipe data |
|-----------|-----------------------|---|------------------|
| 8 | ASAL SEKOLAH | Asal Sekolah Mahasiswa | <i>Nominal</i> |
| 9 | IPS SEMESTER 1 | Indeks Prestasi Semester 1 Mahasiswa | Real |
| 9 | IPK SEMESTER 1 | Indeks Prestasi Kumulatif Semester 1 Mahasiswa | Real |
| 10 | IPS SEMESTER 2 | Indeks Prestasi Semester 2 Mahasiswa | Real |
| 10 | IPK SEMESTER 2 | Indeks Prestasi Kumulatif Semester 2 Mahasiswa | Real |
| 11 | IPS SEMESTER ANTARA 1 | Indeks Prestasi Semester Antara 1 Mahasiswa | Real |
| 11 | IPK SEMESTER ANTARA 1 | Indeks Prestasi Kumulatif Semester Antara 1 Mahasiswa | Real |
| 12 | IPS SEMESTER 3 | Indeks Prestasi Semester 3 Mahasiswa | Real |
| 12 | IPK SEMESTER 3 | Indeks Prestasi Kumulatif Semester 3 Mahasiswa | Real |
| 13 | IPS SEMESTER 4 | Indeks Prestasi Semester 4 Mahasiswa | Real |
| 13 | IPK SEMESTER 4 | Indeks Prestasi Kumulatif Semester 4 Mahasiswa | Real |
| 14 | IPS SEMESTER ANTARA 2 | Indeks Prestasi Semester Antara 2 Mahasiswa | Real |
| 14 | IPK SEMESTER ANTARA 2 | Indeks Prestasi Kumulatif Semester Antara 2 Mahasiswa | Real |
| 15 | IPS SEMESTER 5 | Indeks Prestasi Semester 5 Mahasiswa | Real |

| No | Attribut | Keterangan | Tipe data |
|-----------|--|--|------------------|
| 15 | IPK SEMESTER 5 | Indeks Prestasi Kumulatif Semester 5 Mahasiswa | Real |
| 16 | IPS SEMESTER 6 | Indeks Prestasi Semester 6 Mahasiswa | Real |
| 16 | IPK SEMESTER 6 | Indeks Prestasi Kumulatif Semester 6 Mahasiswa | Real |
| 17 | IPS SEMESTER ANTARA 3 | Indeks Prestasi Semester Antara 3 Mahasiswa | Real |
| 17 | IPK SEMESTER ANTARA 3 | Indeks Prestasi Kumulatif Semester Antara 3 Mahasiswa | Real |
| 18 | IPS SEMESTER 7 | Indeks Prestasi Semester 7 Mahasiswa | Real |
| 18 | IPK SEMESTER 7 | Indeks Prestasi Kumulatif Semester 7 Mahasiswa | Real |
| 19 | IPK LULUS | Indeks Prestasi Kumulatif Lulus Mahasiswa | Real |
| 20 | RANGE IPK | Jarak Indeks Prestasi Kumulatif Mahasiswa | <i>Nominal</i> |
| 21 | FAKULTAS | Fakultas Mahasiswa | <i>Nominal</i> |
| 22 | TAHUN YUDISIUM | Tahun Yudisium Mahasiswa | Real |
| 23 | LAMA KULIAH HINGGA LULUS (TAHUN) | Lama Kuliah Mahasiswa Hingga Lulus Dengan Format Tahun | Real |
| 24 | KATEGORI LULUSAN | Kategori Lulusan Apakah Tepat Waktu(4 tahun, dan 3,5 tahun kuliah untuk S1), atau Tidak Tepat Waktu | <i>Nominal</i> |

Tabel 3.1. menjelaskan tentang tabel Attribut, keterangan, dan tipe data yang digunakan pada penelitian ini.

Validasi data sudah disetujui oleh BIA, dan ketua program studi Sistem Informasi yaitu Ibu Ririn Ikana Desanti, surat dapat dilihat di lampiran.

3.3. Variabel Dependen / Variable Terikat

Variabel dependen adalah variabel yang dipengaruhi oleh variabel independen, data yang didapat mahasiswa/mahasiswi yang sudah lulus pada tahun 2018 hingga 2020 saja, pada *dataset*, variabel dependennya adalah sebagai berikut: KATEGORI LULUSAN.

3.4. Variabel Independen / Variabel Bebas

Variabel Independen adalah variabel yang bisa mempengaruhi antar variabel lain. Pada *dataset*, variabel independennya adalah sebagai berikut: ANGKATAN, ASAL SEKOLAH, PRODI, FAKULTAS, JENIS KELAMIN, IPS SEMESTER 1, IPS SEMESTER 2, SEMESTER ANTARA 1, IPS SEMESTER 3, IPS SEMESTER 4, IPS SEMESTER ANTARA 2, IPS SEMESTER 5, IPS SEMESTER 6, IPS SEMESTER ANTARA 3, IPS SEMESTER 7, IPK SEMESTER 1, IPK SEMESTER 2, SEMESTER ANTARA 1, IPK SEMESTER 3, IPK SEMESTER 4, IPK SEMESTER ANTARA 2, IPK SEMESTER 5, IPK SEMESTER 6, IPK SEMESTER ANTARA 3, IPK SEMESTER 7, TEMPAT LAHIR.

Untuk angkatan dari data yang di dapat dengan *range* 2010-2017, khusus prodi perhotelan angkatan 2017 sudah lulus di 2020 karena bisa wisuda dan lulus di tahun ketiga.

3.5. Metode Penelitian

Metode penelitian adalah merupakan *Quantitative Research*, karena mengukur sebuah masalah data melalui angka, dan juga mengukur sebagai kata-kata deskripsi lulus tepat waktu, atau tidak. Data bisa diubah dalam bentuk statistik dan dijadikan pertimbangan dalam membuat sebuah solusi dan metode yang digunakan bukan dari kuisisioner, survei, *polling*, maupun wawancara yang bersifat pertanyaan, jumlah *participant* pada metode *quantitative* cenderung lebih banyak dibandingkan *qualitative* [31].

Tabel 3.2. Perbandingan Algoritma

| Perbandingan Algoritma | | |
|-------------------------------|---|--|
| Algoritma | Kelebihan | Kekurangan |
| <i>Naïve Bayes</i> | Tidak harus <i>numerical</i> untuk semua variabel yang diprediksi seperti <i>neural network</i> , Bisa dipakai untuk data kuantitatif maupun kualitatif, Tidak memerlukan jumlah data yang banyak, Jika digunakan dalam bahasa pemrograman, code-nya sederhana Bisa digunakan untuk klasifikasi masalah | Hanya dapat menganalisa berbentuk huruf saja, bukan gambar |

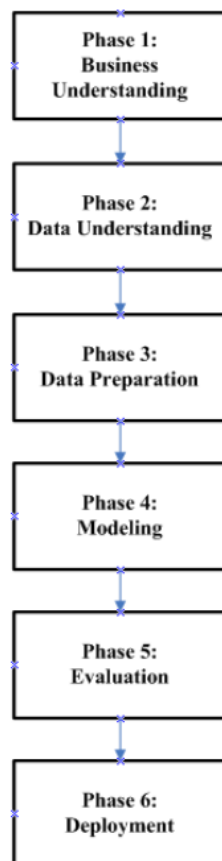
| Perbandingan Algoritma | | |
|-------------------------------|---|--|
| Algoritma | Kelebihan | Kekurangan |
| | biner ataupun <i>multiclass</i> , dibandingkan dengan logistik regresi yang harus variabel dependennya biner ya atau tidak, umumnya semakin banyak variabel makin tepat | |
| Regresi Linear | Bisa untuk <i>label</i> biner maupun <i>multiclass</i> , dalam memprediksi suatu data | data yang digunakan semua harus berbentuk numerikal |
| Jaringan Saraf Tiruan | Dapat meningkatkan akursi dengan cara mengubah <i>hidden layer</i> , bisa memprediksi dan mengklasifikasikan data berbentuk huruf maupun gambar | harus <i>numerical</i> untuk semua variabel yang diprediksi seperti <i>neural network</i> |
| Regresi Logistik Biner | Variabel bebas atau atribut pada regresi logistik tidak harus numerik semua untuk memprediksi variabel dependen, menggunakan logistik logaritma atau logaritmik, cocok untuk regresi logistik 2 pilihan atau <i>true, false</i> . | logistik regresi yang harus variabel dependennya biner ya atau tidak, kecil akurasi bila menebak angka |

Sumber: [16]–[18]

Tabel 3.2. menjelaskan tentang alasan memilih *Naïve Bayes* karena, Tidak harus *numerical* untuk semua variabel yang diprediksi seperti *neural network*, Bisa dipakai untuk data kuantitatif maupun kualitatif, Tidak memerlukan jumlah data yang banyak, Jika digunakan dalam bahasa pemrograman, code-nya sederhana Bisa digunakan untuk klasifikasi masalah biner ataupun *multiclass*, dibandingkan dengan logistik

regresi yang harus variabel dependennya biner ya atau tidak, semakin banyak variabel makin tepat. Alasan memilih Regresi Logistik Biner karena Variabel bebas atau attribut pada regresi logistik tidak harus numerik semua untuk memprediksi variabel dependen, menggunakan logistik logaritma atau logaritmik, cocok untuk regresi logistik 2 pilihan atau *true*, *false*.

3.6. Alur Penelitian



Gambar 3.1. Grafik *CRISP-DM*

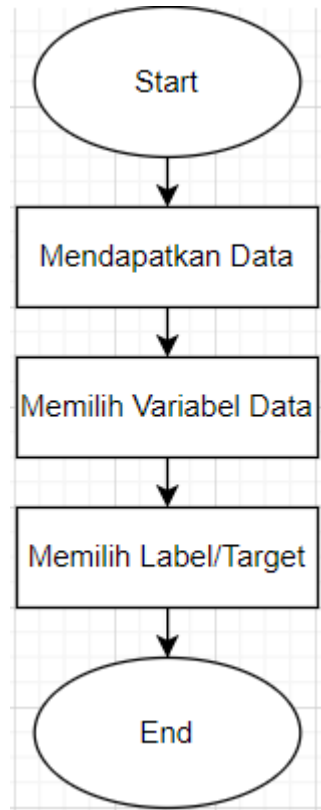
Sumber:[20]

Gambar 3.1. menjelaskan tentang CRISP-DM, CRISP-DM adalah singkatan dari Cross-Industry Standard Process for data Mining yang dikembangkan pada tahun 1996 oleh beberapa analis dari industry, yaitu NCR, SPSS, Daimler Chrysler. CRISP-DM adalah standarisasi proses pada data mining untuk penyelesaian masalah secara umum untuk bisnis atau untuk penelitian.

3.6.1. Business Understanding

Business Understanding pada skripsi ini adalah analisis kelulusan mahasiswa UMN per prodi, dan per semester, tingkat ketidaklulusan tepat waktu masih besar dan meningkat di periode 2018-2020 di UMN, pada bab 4 nanti akan membahas tentang kelulusan yang ada di panduan akademik UMN.

3.6.2. Data Understanding

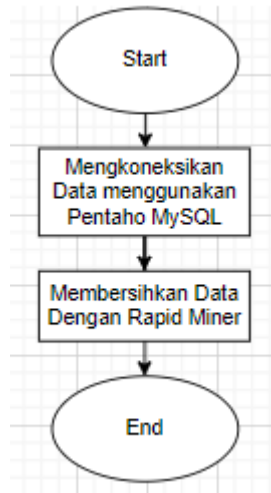


Gambar 3.2. Diagram Alir 1

Sumber:[8]

Pada Gambar 3.2. menjelaskan tentang diagram alir, Proses awal adalah mendapatkan data, lalu menentukan variabel bebas, dan terikat, terakhir menentukan attribute label/target.

3.6.3. Data Preperation



Gambar 3.3. Diagram Alir 2

Sumber: [8]

Gambar 3.3. menjelaskan tentang diagram alir, dengan data yang akan diolah, harus *clean* dan bebas dari *missing value*, maka dari itu dengan menghilangkan *row label* yang *missing* dan *row attribute* yang *missing*, dan *remove duplicates* agar data yang dihasilkan semakin *valid*, mengkoneksikan data ke *Pentaho MySQL*.

3.6.4. Modeling



Gambar 3.4. Diagram Alir 3

Sumber:[8]

Gambar 3.4. menjelaskan tentang diagram alir, *Modeling* yang digunakan adalah *classification*, dengan menggunakan *algoritma data mining Naïve Bayes*, dan *logistic regression*, untuk *data mining* menggunakan *tools rapidminer*, pada penelitian ini dengan membuat kombinasi naive bayes dan regresi logistik dengan model normal, *Cross Validation*, dan mengurangi attribute IPS per semester dengan data *Training* 70%, dan data testing 30%,

setelah mendapatkan kesimpulan, lalu dibentuklah data visualisasi dengan tools *Tableau*.

Tabel 3.3. Perbandingan Software Data Mining

| Software | Kelebihan | Kekurangan |
|-------------|--|---|
| Rapid Miner | <i>Open-source</i> menggunakan 10,000 data Rows, Penggunaan lebih <i>simple</i> karena menggunakan metode <i>drag and drop</i> untuk proses data <i>mining</i> maupun data <i>cleansing</i> | Untuk proses data <i>mining</i> tidak dijabarkan dengan detil, biasanya hanya <i>simple</i> saja contohnya berapa besar akurasi langsung di <i>result</i> |
| Python | <i>Open-source, Multiplatforms</i> ada sistem operasi windows, Linux, mac, <i>neural network</i> di <i>rapid miner</i> bila mau mendapatkan akurasi <i>labelnya</i> harus teks, jadi pakai <i>python</i> | <i>Coding</i> lebih <i>complex</i> dibandingkan rapid miner dan r studio |
| R Studio | <i>Open-source, Multiplatforms</i> ada sistem operasi windows, Linux, mac, fasilitas grafik yang relative baik | <i>Coding</i> lebih <i>complex</i> dibandingkan rapid miner dan hampir sama <i>complex</i> dengan r studio |

Sumber: [16]–[18], [32]–[35]

Berdasarkan Tabel 3.3. menjelaskan tentang pemilihan *software* data *mining* untuk penelitian dari kelebihan dan kekurangan yang ada, maka digunakanlah *rapid miner* karena *Open-source* menggunakan 10,000 data rows, Penggunaan lebih

simple karena menggunakan metode *drag and drop* untuk proses data *mining* maupun data *cleansing*.

Tabel 3.4. Perbandingan Software Data Visualisasi

| Software | Kelebihan | Kekurangan |
|-----------------|--|--|
| <i>Tableau</i> | Pilihan visual yang interaktif, ada grafik bergerak, <i>User friendly</i> , tidak butuh <i>hard coding</i> yang banyak, <i>Dashboard mobile friendly</i> , bisa mengolah data pada <i>mobile</i> , dapat terhubung dengan <i>database</i> yang banyak jenisnya, terdapat fitur <i>story</i> , dan <i>dashboard</i> | Software berbayar, biaya yang cukup mahal |
| <i>Power BI</i> | <i>Open-source</i> software, Bisa membuat <i>dashboard</i> , <i>User friendly</i> , tidak butuh <i>hard coding</i> yang banyak | Jumlah data telah mencapai 2GB, harus <i>upgrade</i> ke versi <i>power BI paid</i> |

Sumber: [36]–[39]

Tabel 3.4. menjelaskan tentang pemilihan *software* data visualisasi untuk penelitian dari kelebihan dan kekurangan yang ada, dengan memilih *software Tableau* karena pilihan *visual* yang interaktif, ada grafik bergerak, *User friendly*, tidak butuh *hard coding* yang banyak, *Dashboard mobile friendly*, bisa mengolah data pada *mobile*, dapat terhubung dengan *database* yang banyak jenisnya, terdapat fitur *story*, dan *dashboard*.

3.6.5. Evaluation

Evaluasi pada penelitian ini dengan perbandingan akurasi model *naive bayes* dan regresi logistik dengan model normal, *Cross Validation*, dan mengurangi *attribute* IPS per semester dengan data *Training* 70% dan data *testing* 30%.

3.6.6. Deployment

Pada penelitian ini, digunakan hanya untuk pembelajaran saja, tahap deployment tidak digunakan karena tidak menggunakan *system* model ke mahasiswa/mahasiswi UMN.